

## **Chapter Two: Contents**

**(Population – 15 October 2001 – LA-UR 01-5712 – Portland Study Reports)**

<b>1. OVERVIEW .....</b>	<b>1</b>
<b>2. SOURCE DATA .....</b>	<b>2</b>
<b>3. GENERATING THE POPULATION.....</b>	<b>5</b>

## Chapter Two—Population

### 1. OVERVIEW

TRANSIMS (Transportation Analysis and Simulation System) is an integrated system of travel forecasting models designed to give transportation planners accurate, complete information on traffic impacts, congestion, and pollution. The Population Synthesizer module constructs a regional population imitation with demographics closely matching the real population. Households are distributed spatially to approximate regional population distribution. The synthetic population's demographics form the basis for individual and household activities requiring travel, and their household locations determine some of the travel origins and destinations. This report outlines how we have constructed the synthetic population for our Portland, Oregon case study. It also briefly summarizes the characteristics of the data and how we verified that the data were correctly generated.

## 2. SOURCE DATA

Fig. 1 illustrates the data flow related to the TRANSIMS Population Synthesizer module. Input data for the module has been collected from a variety of sources: the TRANSIMS Population Synthesizer and Network user manuals provide detailed information on the organization of these input files—here we will only discuss where we obtained the data.

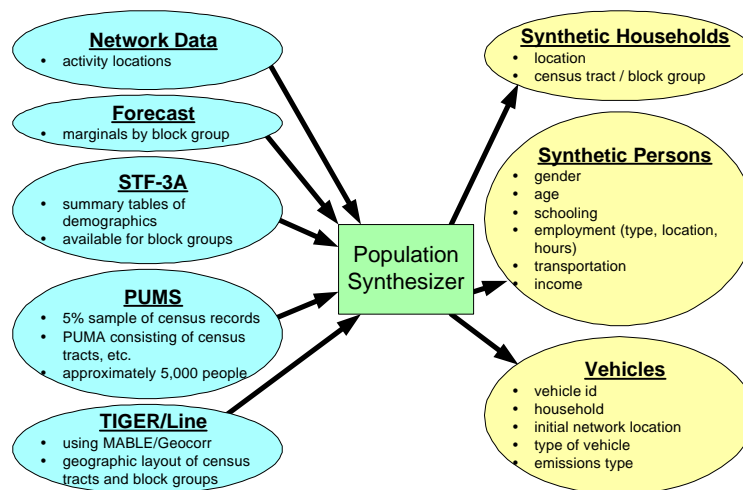


Fig. 1. Data flow for the Population Synthesizer module.

The raw network data were provided by Portland Metro, processed by Los Alamos National Laboratory (LANL), and reviewed by Portland Metro; only activity location data are used during population synthesis. For this, we used the density of residences along streets computed from Portland tax lot data.

Portland Metro provided a 1996 “forecast” (relative to the 1990 census data available for the case study) representing the marginal distribution of the following demographic variables over the 1107 block groups in the study:

- household size
- age of head of household
- annual household income

Although Portland Metro also provided joint distributions of the variables above, they were not used during the population synthesis—we have used them to check the match between the synthetic population and this forecast, however.

The 1990 STF-3A<sup>1</sup> and PUMS<sup>2</sup> data for the states of Oregon and Washington were obtained directly from the U.S. Census Bureau. We extracted the following STF-3A data fields and the corresponding PUMS fields from these databases:

- family households
  - age of the householder [P24]
  - family income [P107]
  - number of workers in the family [P112]
  - poverty status  $\times$  race  $\times$  family type  $\times$  presence and age of children [P124A,B]
- non-family households
  - household type and gender [P17]
  - race  $\times$  household type  $\times$  presence and age of children [P20]
  - age of non-family householder [P24]
  - non-family household income [P110]
  - poverty status  $\times$  age of householder household type [P127]
- group quarters
  - group quarters [P40]
  - group quarters  $\times$  age [P41]

The MABLE/Geocorr web site<sup>3</sup> provided the abstract of the U.S. Census Bureau TIGER/Line data needed for correlating the overlap between block groups in the STF-3A data and PUMAs in the PUMS data.

Finally, Portland Metro provided distributional data for the of 23 CMEM vehicle types<sup>4</sup>, shown in Fig. 2, determined from Oregon motor vehicle records.

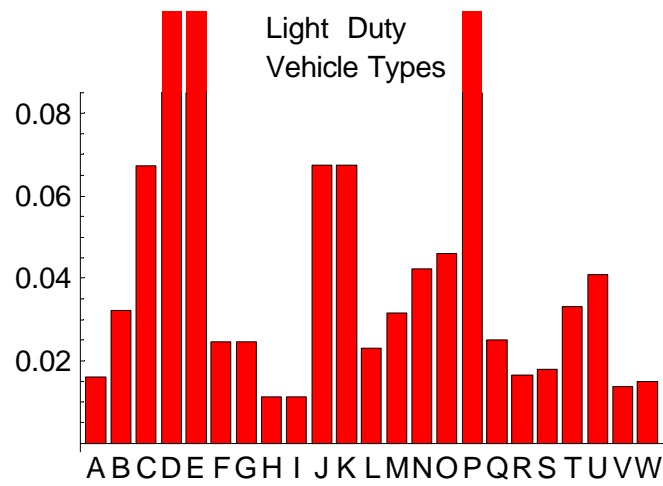
---

<sup>1</sup> Census of Population and Housing, 1990, Summary Tape File 3 on CD-ROM Technical Documentation, prepared by the Bureau of Census.

<sup>2</sup> Census of Population and Housing, 1990, Summary Tape File 3 on CD-ROM Technical Documentation, prepared by the Bureau of Census.

<sup>3</sup> Mable/Geocorr Home Page, <http://plue.sedac.ciesin.org/plue/geocorr>

<sup>4</sup> Barth, M., T. Younglove, T. Wenzel, G. Scora, F. An, M. Ross, and J. Norbeck (1997), "Analysis of Modal Emissions for a Diverse in-use Vehicle Fleet." *Transportation Research Record, No. 1587*, Transportation Research Board, National Academy of Science, pp. 73-84.



*Fig. 2. Distribution of 23 CMEM vehicle types determined from Oregon motor vehicle records.*

### 3. GENERATING THE POPULATION

We generated the synthetic population using the procedure specified in the TRANSIMS Population Synthesizer (see the Population Synthesizer general documentation). Populations for each PUMA of the six PUMAs (01200, 01300, 01400, 01500, 01901, 01902) were generated separately and then merged into a single file, as recommended by the user manual. The contents of the six configuration files used to generate these segments of the population, along with the configuration file used to locate the population and generate the vehicles (*allstr.cfg*) are given in Volume Eight (*Appendix: Scripts, Configuration Files, Special Travel Time Functions*), Chapter Two (*PS-1*).

We performed a variety of tests to verify that the Population Synthesizer operated correctly. The straightforward tests of the software included:

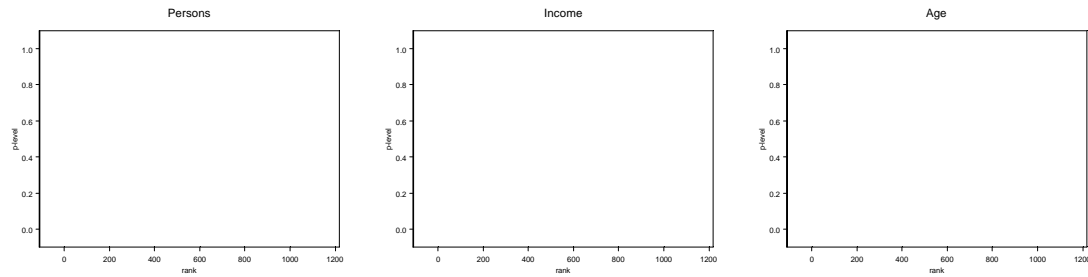
- The correct records are extracted from the PUMS data file.
- The correct fields and records are extracted from the STF-3A database.
- The forecast marginal records are correctly read.
- Demographic variables and block group numbers in the above records are correctly indexed.
- The multiway input weights to the “rake” procedure are constructed correctly.
- The marginals in the multiway output tables of the “rake” procedure match those for each block group of the input constraints to “rake” (this applies to the base “rake” run and the forecast “rake” run).
- The iterative proportional fitting procedure in “rake” finds the correct proportion matrix for each block group.
- The correct number of households is generated for each block group.

We performed tests (such as the chi-squared test) on the output to verify the statistical correctness of the results:

- The marginal distributions of the demographics for each block group match the STF-3A and forecast input data for the base and forecast populations, respectively. The results of this chi-square test are shown in Fig. 3.
- The PUMS records drawn for each block group’s base and forecast population match the probability distribution of the “rake” output multiway table.
- The odds ratios in the multiway table fit to the STF-3A matches that for the PUMS data.

We also inspected the located populations and vehicles to verify that:

- Households are placed at activity locations proportionally to area of residences along the street.
- Households are not placed in commercial, industrial, etc., areas.
- Vehicles are assigned according to household demographics.
- Vehicles are parked at the location nearest the household.



*Fig. 3. Probability levels sorted by rank for a chi-squared test of the marginal demographics in 1107 block groups. The strong linearity of the plots indicates a good match to the forecast demographics.*